DOCUMENT RESUME

ED 216 629 HE 015 082

AUTHOR Blackburn, Robert T.; Conrad, Clifton F.
TITLE Research on Program Quality.
PUB DATE 19 Mar 82
NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, March 19, 1982).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *College Programs; *Educational Quality; *Evaluation Criteria; Graduate Study; Higher Education; Measurement Techniques; Peer Evaluation; Predictor Variables; Professional Education; *Program Evaluation; *Reputation; Research Methodology; *Research Problems; Standards; Undergraduate Study

ABSTRACT

Program evaluation research, including reputational studies and studies based on objective indicators, and research on the quantitative correlates of program quality are reviewed. Reputational studies, which have dominated research on program quality in higher education, emphasize peer evaluation. Some major reputational studies of graduate programs since 1924 are briefly described. It is suggested that reputational studies based largely on faculty rating of faculty quality have yielded highly stable ratios of top-ranked graduate programs and institutions. Reputational studies of professional programs, which are relatively new, are also considered, along with recent reputational studies of undergraduate programs. One criticism of reputational studies is rater bias. Other criticisms include the fact that reputational studies yield more a rank order of excellence than a position on a normative scale. Quantitative studies of program quality encompass criteria and objective indicators, and an index of the indicators, which is used to rate programs. Major criteria and widely-used objective indicators that have been used to evaluate program quality are listed. Additionally, researchers have attempted to identify objective program characteristics, correlates of quality, that are associated with programs of high quality. These studies have focused almost exclusively on the graduate domain. Criticisms of studies based on objective indicators and of studies of correlates of quality are also examined. A bibliography is appended. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED216629

RESEARCH ON PROGRAM QUALITY

Robert T. Blackburn
The University of Michigan

Clifton F. Conrad
The University of Arizona

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Clifton F. Conrad

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Paper presented at the Annual Meeting of the American Educational Research Association, New York, March 19, 1982.

HE 015 082

Program quality in higher education--what it is, how to identify it, and how to foster it--has been an enduring concern among educators since the founding of Harvard College in 1636. In the twentieth century alone, concerns about quality can be traced to the widespread adoption of accreditation procedures, the Flexner (1910) report on medical schools, and the first major comparative ranking of institutions (Hughes, 1925).

Today, a variety of social forces make assessing program quality a high priority problem. Public disenchantment with educational institutions, pressures for accountability, and declining resources are but three factors receiving attention. State systems must decide which programs to retain in the face of declining enrollments and limited resources. Private liberal arts colleges, especially those with essentially open admission policies, need to demonstrate for their very survival that their high costs are linked to a quality education. Local tax revolts are forcing community colleges to decide which programs to phase out. And the government and foundations have to identify quality programs before they invest their resources to solve social problems.

These practical reasons for assessing program quality, coupled with scholarly interest in developing new knowledge about quality, have stimulated a considerable body of research on the subject. Research on program quality has centered around two major approaches: evaluation studies and studies of the quantitative correlates of quality. Program evaluation studies, which have been conducted both at the graduate and undergraduate level, have aimed at ranking departments, professional schools, and occasionally, entire institutions. Program quality has been evaluated either through the "reputational" method or on the basis of "objective" indicators. The second major approach to research on quality has been directed not to the evaluation of program quality

per se, but to the identification of quantifiable program characteristics (such as program size) that are associated with programs considered to be of high quality.

The purpose of this paper is to review and critically examine these two threads of research and to analyze the implications of these efforts for current and future research. The paper is divided into two major sections. The first section examines program evaluation research (including both "reputational" studies and studies based on "objective" indicators), and the second looks at research on the quantitative correlates of program quality. In each section the pertinent literature and major findings are examined, the major evaluative criteria and methodological procedures used in research studies are reviewed, and a critique of their limitations and strengths provides the basis for a number of methodological and substantive recommendations concerning future research on program quality.

Evaluation Studies

Reputational Studies

Reputational studies have long dominated research on program quality in higher education. Built upon a foundation of peer evaluation, these studies have been criticized as severely by some as they have been heralded by others. Despite the controversy surrounding the approach, much of the best-known research on program quality has been in this tradition.

The most salient characteristic of reputational studies is the emphasis placed on peer evaluation--a practice with deep roots in the history of the professions. In reputational studies, judgments of program quality are made by faculty raters and/or by other experts such as department chairmen, deans, and professional association and agency officers. While the particular peer raters usually differ across studies, all studies share the common assumption that it is

experts in the field who are best qualified to make assessments of program quality. Further, reputational studies follow a similar procedure: The researcher selects one or more criteria (such as faculty quality) to serve as a basis for evaluation, employs a panel of experts to rate programs in terms of that criteria, and then combines individual panelists' responses in order to generate a ranking or rating of programs by institution. Since most reputational studies have evaluated programs at the doctoral level, we turn first to a review of studies at the graduate level before examining the literature on professional schools and undergraduate programs.

The first major reputational study of graduate programs was conducted by Hughes in 1924. Using panels of scholars in each discipline to identify top scholars in the field, Hughes (1925) ranked thirty-eight universities in twenty graduate fields according to the number of top scholars they employed. A second Hughes (1934) study rated fifty-nine universities in thirty-five fields according to judgments of staff and facilities made by faculty experts. As noted by Lawrence and Green (1980:4), the two Hughes studies established important precedents for later studies of program quality: a focus on the graduate (doctoral) level, reliance on ratings by academics rather than outside observers, and an emphasis on the nation's most visible institutions. No less important, and knowingly or otherwise, Hughes set the precedent that the major criterion for evaluating quality is to be faculty quality.

A quarter of a century after the second Hughes study, Keniston (1959) selected department chairmen from institutional members of the American Association of Universities to rank twenty-four graduate programs at the twenty-five institutions leading in doctorate production. In addition to ranking programs by institution, Keniston rank-ordered the top twenty institutions. His study, as well as the two earlier Hughes studies, was severely criticized by researchers on

methodological grounds. These criticisms, which included complaints about geographical and rater bias, led to subsequent reputational studies aimed at updating and improving ratings of graduate programs.

Probably the most methodologically sophisticated of the reputational studies were designed under the auspices of the American Council of Education (ACE), which sponsored ratings of doctoral programs in the arts and sciences in 1964 and 1969 (Cartter, 1966; Roose and Anderson, 1970). The Cartter study yielded institutional rankings in twenty-nine disciplines at a total of 106 universities, while the Roose-Anderson study ranked programs in thirty-six disciplines at 130 universities.

In both ACE surveys, ratings within each discipline were made by department chairmen, distinguished senior scholars, and selected junior scholars. Each respondent rated doctoral programs in his own field in terms of two criteria: quality of the graduate faculty and effectiveness of the graduate program. The individual ratings were then averaged and the leading programs in each field were rank-ordered in two separate lists, one based on faculty quality and the other on program effectiveness. The two sets of rankings in both the Cartter and Roose-Anderson studies corresponded closely.

Several scholars have rightly noted that the methodological procedures in the ACE studies gave careful attention to the statistical validity and reliability of the ratings (Clark, 1976; Lawrence and Green, 1980). Cartter (1966) used detailed analyses to show that the rankings were similar whether respondents were separated by rank, institution of employment, degree, or geographical region. Moreover, he showed consistency between the ACE ratings and ratings by smaller and more select panels in several fields.

Since 1970, only two major reputational studies of doctoral program quality have been published. One study, conducted under the joint sponsorship of the Council of Graduate Schools (CGS) and the Educational Testing Service (ETS), used

faculty raters to assess and rank graduate programs (Hartnett, Clark, and Baird, 1978). As in the ACE ratings, programs were rated in terms of faculty quality and program effectiveness. Another study, by Ladd and Lipset, asked more than 4,000 faculty members at four-year colleges and universities to "name the five departments nationally in your discipline that have the most distinguished faculties" (Scully, 1979:7). Ladd and Lipset then ranked departments that were rated among the top five by at least 10 percent of the respondents in each of nineteen fields.

Comparison of the rankings of graduate programs across all of the studies discussed above reveals a consistent pattern of findings: Although there are minor variations in the absolute rankings of programs, reputational assessments of graduate program quality have consistently identified the same graduate programs at the top of the rankings.[1] In terms of departments with middle and lower rankings, the absolute rankings vary considerably but still show a high degree of stability across broad groups of departments.

Although most reputational researchers discourage the aggregation of departmental rankings into institutional rankings, many have felt compelled to make such aggregations (American Educational Council, 1971; Magoun, 1966; Morgan, Kearney, and Regens, 1976; National Science Board, 1969; Petrowski, Brown and Duffy, 1973). When universities with high overall standings in the recent Ladd-Lipset and ACE surveys are compared with the overall rankings of Hughes (1925) and Keniston (1959), the stability of university reputations is clearly seen: All but one of the top twenty institutions are still at the top, and less than a handful of new ones have been added. The remarkable stability in the rankings is supported by the fact that six institutions

---

[1] Although the CGS-ETS study does not identify institutions, the authors' report that their rankings are very similar to the 1966 and 1970 ACE lists (Hartnett, Clark, and Baird, 1978). The correlations across time are all above 0.9.

(California-Berkeley, Chicago, Harvard, Michigan, Princeton, and Yale) have been ranked among the top ten institutions in all five of the major reputational studies of graduate education (Cartter, 1966; Hughes, 1925; Keniston, 1959; Ladd-Lipset in Scully, 1979; Roose and Anderson, 1970).

In summary, the reputational method has served as the major way of evaluating program quality at the graduate level. Reputational studies, which have placed major emphasis on faculty quality as judged primarily by faculty raters, have yielded highly stable ratings of top-ranked graduate programs and institutions.

While most of the well-publicized studies of program quality have examined graduate departments in the arts and sciences, recent interest in professional programs has led a few researchers to apply the reputational method to assessing professional program quality. Two widely-publicized studies evaluated programs in at least three professional fields, and several other studies (conducted both by academics and by the professions themselves) have focused on a single field.

The first major study of professional education, completed by Margulies and Blau (1973), rated programs in seventeen professional fields. Deans of professional schools served as raters and were asked to identify the top five programs in their field. The responses were then aggregated and programs were ranked, by institution, for all seventeen fields. Criticized severely for the low overall response rate to their survey, Blau and Margulies subsequently completed a replication study using the same fields plus one more. Even though the response rate increased dramatically, the rankings remained about the same in all fields (Blau and Margulies, 1974-1975).

A second major study of professional programs rated programs in law, education, and business (Cartter and Solomon, 1977). Unlike the earlier study by

Blau and Margulies, Cartter and Solomon used faculty members as well as deans as raters; had raters evaluate faculty and program attractiveness on Likert-type scales rather than simply select the top five fields; and provided raters with a list of institutions to evaluate rather than asking them to name, by recall, top programs in their field. Despite these methodological differences, comparisons between the Blau-Margulies and Cartter-Solomon rankings (in law, education, and business) reveal many similarities. There are, however, some notable differences, which may be due primarily to differences in sample size and selection and in the survey instruments used (Munson and Nelson, 1977).

Other reputational studies of professional education have ranked programs in a single field. For example, Cole and Lipton (1977) ranked top medical schools; Carpenter and Carpenter (1970) ranked library science programs; MBA magazine (1974; 1975) ranked the top fifteen business programs in two reputational studies; and Juris Doctor magazine (1976:18) ranked the top twenty law schools.

Although reputational studies of professional programs are relatively new, two important observations can be made about these studies. First, while they have been influenced by earlier reputational studies at the graduate level (particularly the methodology of the 1966 Cartter study), they often have utilized more diverse groups of raters and more diverse rating criteria. Second, as Lawrence and Green (1980:22) have reported, most reputational rankings have consistently identified the same professional schools at the top, although there is some variance in the ranks of institutions across studies. Three rankings of law schools (Blau and Margulies, 1974-1975; Cartter and Solomon, 1977; Juris Doctor, 1976) have in common seven institutions among those at the top of the rankings; three rankings of business schools (Blau and Margulies, 1974-1975; Cartter and Solomon, 1977; MBA, 1975) share six

institutions at the "top," and two rankings of medical schools (Blau and Margulies, 1974-1975; Cole and Lipton, 1977) have eight institutions in common.

In the last few years, there have been only two reputational studies of program quality at the undergraduate level. One study, sponsored by Change magazine, identified ten "leading" institutions in terms of national influence (Johnson, 1978). The list of ten institutions is very similar to the top-rated institutions as identified in the aforementioned reputational studies of graduate programs (cf. Petrowski, Brown, and Duffy, 1973).

A second, more recent study (Solomon and Astin, 1981; Astin and Solomon, 1981) represents the most ambitious attempt to rate undergraduate programs using a reputational approach. Solomon and Astin asked faculty raters to rank departments in their field of specialization (from a list of about 100 institutions) on the basis of six rating criteria. After faculty in seven departmental specializations had ranked departments according to each of the six rating criteria, the researchers identified the ten top-ranked institutions on each list. The final list of top-ranked undergraduate departments in each of the seven fields included those departments which had been ranked in the top ten according to at least one criterion.

While the Solomon-Astin lists of top-ranked institutions includes most of the institutions identified in earlier reputational studies of graduate education, the researchers identified a fairly large number of institutions that have not appeared in previous reputational studies of top-ranked institutions. This finding is especially significant, for it suggests that broadening rating criteria to include dimensions such as "faculty commitment to undergraduate teaching" leads to the identification of programs of excellence which are otherwise ignored when the traditional criterion of faculty quality (pre-

sumably a judgment of faculty scholarly achievement) used in reputational studies serves as the main criterion for evaluation.

Astin and Solomon (1981) admit that their reputational study is only preliminary, given the relatively small samples of departments and institutions included in their study. Still, the researchers have extended the assessment of quality beyond faculty quality and found that diversifying rating criteria (at least at the undergraduate level) is likely to lead to the identification of "quality" programs that would otherwise be overlooked.

Critique

Given the visibility of most of the reputational studies, it is not surprising that they have received a steady stream of criticism. Most of this criticism has been methodological rather than substantive, and some of it has been uneven, unsupported, and unfair.[2] In this critique, our intent is to examine the most valid criticisms and major strengths of reputational studies, as well as to make some suggestions regarding future research. We begin by discussing two methodological weaknesses of reputational studies.

First, reputational studies can be justly criticized for rater bias, which can take several forms. For example, the overall reputation of an institution may influence raters' assessments of a particular department, expecially when raters are not well-informed. Such "halo effects" have been found by researchers at both the undergraduate and graduate level (cf. Solomon and Astin, 1981:27). Also, departmental or institutional reputation may "lag" several years behind

---

[2]For example, one argument advanced against reputational studies is that they are intrinsically "subjective." This criticism is grossly unfair, for it ignores the fact that reputational studies are purposively based on "subjective" peer evaluations. Moreover, subjectivity cannot be completely avoided in evaluation studies, regardless of the evaluation technique used. As Cartter noted, even so-called "objective" measures (such as number of Nobel laureates on the faculty) are, for the most part," 'subjective' measures once removed" (1966:4).

current quality, thereby introducing another source of rater bias (Hughes, 1925; Cox and Catt, 1977). In addition, in most reputational studies, scholars from the highly ranked universities are over-represented as raters, and may tend to rate higher those departments with similar emphases as their own (Webster, 1981:21). Other sources of rater bias--such as "alumni effects" (the tendency of raters to highly rate their alma maters) and the influence of an institution's age or size on raters' perceptions of quality--have been noted in the literature (Lawrence and Green, 1980:10). However, there is insufficient evidence to conclude that these sources affect raters' assessments of program quality.

A second, and related, methodological criticism of reputational studies is that many raters are not sufficiently well-informed to make judgments about the quality of programs at other institutions, particularly when the evaluative criteria extends beyond faculty quality. Since most reputational studies have ranked programs presumably on the criterion of faculty scholarly achievement, this criticism often has been rejected on the grounds that who but faculty are most competent to judge their peers (Blackburn and Lingenfelter, 1973:24). While this rejoinder has some merit, it overlooks the fact that many faculty are still relatively uninformed about faculty at other institutions, particularly at less visible universities. Moreover, if reputational studies are to evaluate something more than faculty quality, raters should be selected who are well-informed about programs in terms of all the criteria serving as the basis for evaluation.

In addition to these methodological criticisms, five substantive criticisms can be directed against most reputational studies. First, because the emphasis is primarily on ranking programs rather than evaluating them against a normative standard of quality, reputational studies yield more a rank order of excellence than a position on a normative scale. Put differently, reputa-

tional studies do not inform us whether a particular program is of high, medium or low quality; they tell us only its relative quality vis-a-vis those programs to which it is compared.

Second, the criteria used to assess quality in most reputational studies are severely limited. Studies at the graduate level have relied on a single yardstick to assess quality: faculty quality.[3] As a consequence, most reputational studies are peer judgments of departmental faculty quality, presumably based almost entirely on scholarly publication. Given the use of a single criterion in reputational studies of graduate programs, it is hardly surprising that the same institutions are consistently ranked at the top.

By reifying the criterion of faculty quality, most reputational studies have ignored a number of other potential evaluative criteria, most notably teaching and program effectiveness. As stated in one review of reputational studies at the graduate level: "They say little or nothing about the quality of instruction, the degree of civility or humaneness, the degree to which scholarly excitement is nurtured by student-faculty interactions, and so on" (Hartnett, Clark, and Baird, 1978). In our view, faculty quality alone is an insufficient criterion for assessing overall program effectiveness--whether at the graduate, professional, or undergraduate level.[4]

---

[3] While several of the reputational studies have used more than one criterion, faculty quality has been the major criterion in all studies at the graduate level. For example, the Cartter (1966) study of doctoral programs used "effectiveness of the doctoral program" as well as "quality of the faculty." However, the high correlation between the two sets of rankings in the Cartter study suggests that raters probably chose to emphasize what they knew best (i.e., quality of the faculty) when evaluating "effectiveness" (Blackburn and Lingenfelter, 1973:24). In effect, then, faculty quality was the major criterion.

[4] It should be noted that expanding evaluative criteria beyond faculty quality may lead to different rankings of programs. Several evaluation studies, one conducted at the graduate level and the other at the undergraduate level, found that when independent assessments of commitment to teaching/program effectiveness and faculty quality are made, programs are ranked quite differently (Astin and Solomon, 1981; Hartnett, Clark, and Baird, 1978).

Third, because most reputational studies have used the department as the unit of analysis, the institutional environment outside of the department is rarely considered in the evaluation. Especially at the undergraduate level, where students are exposed to a variety of fields and extra-curricular influences, that quality assessment should take the internal, institutional environment into account is a persuasive argument. Indeed, the argument has been forcefully made in terms of the graduate level: ". . . quality is a property of the total institution . . . (and) a component of high quality can be expected to flourish best within an environment of high quality" (National Science Board, 1969:50).

Fourth, since they have focused exclusively on ranking the "top" 20 to 150 institutions, all reputational studies have ignored programs at most of the postsecondary institutions in the nation. As a result, we know remarkably little about program quality throughout most of higher education.

Fifth, while reputational studies at the graduate level have received considerable attention, little attention has been given to professional schools and undergraduate programs. To be sure, there have been a substantial number of studies of professional schools in recent years. But with only two exceptions (Blau and Margulies, 1974-75; Cartter and Solomon, 1977), all of these studies have been limited to a single field. At the undergraduate level, there have been only two major studies of program quality. One of these studies (Solomon and Astin, 1981) is a sophisticated reputational study. Yet, as noted earlier, even Solomon and Astin admit that further research needs to be done at the undergraduate level.

Despite all of the above criticisms, it would be unfair to conclude that reputational studies have not contributed to the understanding of program quality in American higher education. Reputational studies at the graduate

level have provided defensible ratings of "leading" departments in terms of faculty scholarship (Blackburn and Lingenfelter, 1973:25). At the professional school level, reputational studies have consistently identified most of the same "top" programs while using multiple criteria for evaluation. And, at the undergraduate level, progress has been made in assessing program quality based on multiple criteria.

Over a decade ago, the controversy over the last of the ACE-sponsored reputational studies (Roose and Anderson, 1970) prompted the American Council on Education to discontinue sponsorship of any further reputational studies. Since then, most reputational studies have been conducted at the professional school and undergraduate level. In our view, controversy over existing studies should not deter researchers from conducting future reputational studies. If future reputational studies are designed to respond to the criticisms raised here, we are confident that they can make a more important contribution to evaluating quality in higher education. Especially if methodological improvements are made, if quality is evaluated through multiple criteria, if normative standards of quality are used, and if quality is evaluated not just at "leading" schools, then future reputational studies can shed considerable light on program quality at the undergraduate and professional as well as the graduate level.

Studies Based on Objective Indicators of Quality

Although reputational studies have dominated research on program quality, criticism of the reputational approach has led some researchers to explore more "objective," quantifiable procedures for assessing program quality in higher education. In a word, researchers have sought to "quantify" quality in order to measure it empirically.

Quantitative studies of program quality generally follow a straightforward

format: the evaluator selects criteria (such as faculty scholarship) along with objective indicators (such as average number of refereed articles per faculty member for each of the last five years) on an a priori basis, develops an index of those indicators, and then rates programs on the basis of that index. In some studies, only one or two indicators have been used to assess quality, but in many cases researchers have used multiple criteria and multiple indicators. As a result, a wide range of criteria and indicators have been used to assess and rank graduate programs and undergraduate colleges and, in a few cases, provide composite institutional rankings (Adams and Krislov, 1978; Bowker, 1964; Calvert, Pitts, and Dorion, 1971; Clemente and Sturgis, 1974; Cox and Catt, 1977; Eells, 1960; Glenn and Villemez, 1970; Hurlbert, 1976; Knudsen and Vaughan, 1969; Krause and Krause, 1970; Lewis, 1968; Siebring, 1969; Somit and Tanenhaus, 1964).

One of the first major quantitative studies of program quality at the graduate level, published by Bowker (1964) nearly two decades ago, illustrates the general procedure followed by researchers using this approach to examining program quality. In order to develop a ranking of U.S. graduate schools in the social sciences-humanities and the sciences, Bowker developed a quantitative index which included objective indicators in each of his two major fields. In the social sciences-humanities, for example, he used four quantifiable indicators of quality: number of former Woodrow Wilson Fellows on the faculty; number of Woodrow Wilson fellows choosing to attend the institution; number of American Council of Learned Societies Award winners on the faculty; and the number of Guggenheim Memorial fellows on the faculty. On the basis of these indicators of quality, along with two separate indicators of quality in the sciences, Bowker developed institutional rankings of graduate programs in the social sciences-humanities and in the sciences (1964:6-10).

Since Bowker's study, most research based on objective indicators of quality has focused on the evaluation and ranking of graduate programs in particular disciplines, with graduate programs in sociology and psychology receiving the most attention. Numerous rankings of graduate departmental quality have been published, but many of these rank different programs and thus are not comparable to one another. In those cases where rankings can be compared, there is a pattern: Although the absolute ranks of graduate programs vary (sometimes considerably, others hardly at all) from one study to another, objective assessments have consistently identified the same graduate programs at the top of the rankings. To be sure, the high degree of consensus across the rankings of graduate departments and institutions may be attributed to heavy reliance on the criterion of faculty research productivity and research-oriented indicators, and further, that researchers have been concerned only with ranking a relatively small number of institutions (usually 20 to 80). Still, the relatively high agreement among the rankings is noteworthy, especially since most of the studies have been based on more than one criterion and multiple, frequently different, indicators of quality.

The debate over rankings of graduate programs and institutions has prompted a number of scholars to compare reputational rankings with rankings based on objective indicators of quality. After ranking graduate programs and institutions based on objective measures, several researchers have gone on to compare their rankings with the ACE-sponsored rankings by Cartter (1966) and Roose-Anderson (1970). A consistent pattern emerges across these comparisons: There is a close correspondence between objective and subjective/reputational rankings, though the absolute ranks of particular programs usually vary somewhat (Adams and Krislov, 1978; Clemente and Sturgis, 1974; Cox and Catt, 1977; Glenn and Villemez, 1970; Hurlbert, 1976; Knudsen and Vaughan, 1969; Lewis,

1968; Siebring, 1969). In other words, the research so far suggests that for graduate programs the same institutions will be ranked at the top whether or not departments are evaluated "objectively" through quantifiable indicators or "subjectively" through peer ratings.

With the exception of a study which ranked law schools on the basis of a "resources index" (Kelso, 1975), there has been little published research evaluating and ranking professional schools on the basis of objective indicators of quality. However, the evaluation and ranking of undergraduate colleges has received a modest amount of attention in the last several decades.

Studies of undergraduate colleges tend to utilize one criterion--student achievement--as the basis for rating undergraduate institutions. Using this criterion, Bowker (1964) ranked the top twenty institutions on the basis of the number of baccalaureate recipients who received Woodrow Wilson Fellowships and the number of baccalaureate recipients who eventually earned a doctorate; Krause and Krause (1970) ranked colleges according to the number of their baccalaureate graduates who contributed articles to Scientific American; Dubé (1974) ranked 100 undergraduate colleges according to the number of their alumni who entered medical school; Tidball and Kristiakowsi (1976) ranked institutions according to the proportions of their baccalaureate graduates who went on to earn doctorates; and Astin and Solomon (1979) ranked institutions on the basis of a selectivity index, an index based on an estimate of the average academic ability of an institution's entering freshmen (Astin and Henson, 1978).

Several other studies have used multiple criteria and multiple indicators as a basis for ranking undergraduate colleges (Brown, 1967; Gourman, 1967; Jordan, 1963). By way of illustration, Brown (1967) grouped colleges on the basis of eight indicators: 1) proportion of faculty with a doctorate; 2) average compensation (salary and fringe benefits) per faculty member;

3) proportion of students going on to graduate school; 4) proportion of graduate students; 5) number of library volumes per full-time student; 6) total number of full-time faculty; 7) student-faculty ratio; and 8) total current income per student.

Comparisons of the published rankings of undergraduate colleges, whether the rankings are based on single or multiple criteria and indicators, reveal substantial disparities. Although many of the same colleges appear on all or most of the rankings, studies using objective assessments have not consistently identified the same colleges at the top of the rankings. Moreover, even in those studies where many of the same institutions have been ranked, there is wide variation in the absolute ranks of colleges. In short, research aimed at ranking undergraduate institutions has not resulted in the high degree of agreement across studies that has characterized research on graduate programs.

Having thus far reviewed studies of both graduate and undergraduate education, we turn now to a more detailed look at the criteria and indicators that have been used in this approach to examining program quality. Following that, we discuss the major strengths and limitations of the research based on these objective indicators.

Objective Indicators of Quality

Consonant with the recent interest in the objective assessment of academic programs has been the ongoing search for indicators of program quality. A few researchers have designed studies to identify criteria and indicators of quality by surveying selected respondents (Barak, 1976; Clark, Hartnett, and Baird, 1976; Clewell, 1980; Fotheringham, 1978; Longanecker, 1978; Lynd, 1976). Most researchers, however, have relied on their own insight and imagination to develop quality indicators or have borrowed indicators from accreditation or fugitive reports on program quality.

Out of this collective search for indicators have emerged several major criteria and a large number of indicators of quality. Some of these objective indicators have been used singly and others have been used in combination; most have been used in evaluating graduate programs, though many have been employed in studies of undergraduate education; and there is considerable controversy among researchers regarding the appropriateness of many of the indicators. Nonetheless, these indicators represent the state-of-the-art in the continuing search for a more objective foundation for assessing program quality. Listed in Appendix A are the major criteria and some of the more widely-used objective indicators that have been employed by researchers using this approach to evaluating program quality in higher education.

Critique

Since studies based on objective indicators have been viewed more as a check on reputational studies than as an independent method of rating programs, there has been relatively little formal scholarly criticism of this thread of research, and most of that criticism has been confined to researchers who have raised objections to specific indicators. However, several overall criticisms--both methodological and substantive--should be raised about evaluation studies that have been based on objective indicators.

Most important, serious methodological questions can be asked about the appropriateness of many of the indicators or measures that have been employed in these studies. Instead of establishing the appropriateness of their measures, either through conducting independent surveys of professionals or by selecting indicators from the literature (cf. Fotheringham, 1978), most researchers apparently have selected indicators on the basis of their personal preferences and, presumably, data availability. As a consequence, many--if not most--of the indices used to assess various evaluative criteria may not be

good indicators of program quality.

As a result of the rather haphazard approach to selecting indicators, there is little consensus in the literature concerning the most appropriate measures of virtually all the criteria that have been used to evaluate program quality. For example, there is little agreement among researchers regarding the best measure(s) of undergraduate student quality. Is it the proportion of graduates who go to graduate school? Proportion of graduates who earn doctorates? Academic ability of freshmen as measured by standardized tests? Of course, differences over the appropriateness of any measure or index are to be expected in any area of investigation. But when there is such little agreement among researchers regarding the best measures, it must be concluded that far more attention should be given to establishing the appropriateness of those indicators that provide the foundation for assessing program quality.

In the search for measures of program quality, researchers are well-advised to seek quality indicators that assess adequacy as well as frequency or volume. Most of the existing research is based on indicators that assess easily quantifiable program characteristics (such as number of volumes in the library) rather than adequacy (such as comprehensiveness and accessibility of the library). Some researchers have attempted to measure the adequacy of certain program characteristics, however. In the area of scholarly productivity, for example, some have sought to separate quantity of publication from quality of publication. Instead of simply measuring scholarly publication using frequency tabulations, they have constructed "citation indexes" as a measure of the quality of scholarship (cf. Smith and Fiedler, 1971). While "citation indexes" represent a serious attempt to improve quality measurement, they have--like other indices that attempt to measure quality--come under attack from various scholars (Webster, 1981:22-23). The debate over virtually all indicators of quality serves to underscore the more telling point: the search for appropriate indicators of program quality is

only in its infancy. Put another way, researchers need to improve the "objective" assessment of the valuative, subjective concept of quality.

Questions about the appropriateness of indicators are linked to a second major criticism that can be made about most studies based on objective indicators, namely, that most of the research is highly "subjective." Since research is infrequently based on "objective" measures that have been established through research, most studies reflect the researcher's biases in the selection of indicators. Ironically, much of the research in this tradition can be fairly criticized for its failure both to acknowledge and reduce the "subjectivism" that is supposed to be eliminated in this approach to evaluation research.

Third, studies at the graduate level have been heavily based on the single criterion of faculty research productivity. Although a range of research-oriented indicators have been used to measure that criterion, the reliance on a single criterion has doubtless contributed to the marked similarity in rankings of programs across comparable studies. Most studies at the graduate level have neglected other evaluative criteria, especially teaching and program effectiveness.

Fourth, most of the indicators used to assess quality are useful only for ranking schools at the very top (Webster, 1981:22). For example, the number of Guggenheim Memorial fellows on the faculty or the number of students contributing articles to Scientific American may be good indicators of faculty and student quality at leading institutions. But for the vast majority of postsecondary institutions, their utility, relevance, and appropriateness are highly questionable.

Fifth, none of the studies using objective indicators have assessed programs in terms of contributions to bringing about changes in students--presumably the major purpose of education at the graduate and professional levels as well as the undergraduate level. This "value added" concept has been trumpeted by many scholars in the higher education literature (Astin, 1977; Bowen, 1977; Blackburn

and Lingenfelter, 1973; Clark, Hartnett, and Baird, 1976; Lawrence and Green, 1980), but such an approach has not been used in any large-scale study of program quality in higher education.

Finally, several criticisms made earlier about reputational studies also can be applied to research based on objective indicators. First, most studies have focused on the department as the unit of analysis and thus have failed to use indicators of quality that take the institutional environment into account. Second, because almost all studies have been concerned with ranking programs, they have identified a rank order of excellence rather than evaluated programs in terms of a normative standard of quality. Third, by making the questionable assumption that programs have identical goals, objective indicator studies have assessed all programs against the same yardstick. Fourth, they have consistently focused on ranking only the "top" institutions, thereby ignoring programs in the vast majority of postsecondary institutions.

Despite all of these criticisms, it is important to point out that studies based on objective indicators have already made a contribution to evaluating program quality in higher education. Methodologically, many studies have gone beyond reputational studies not only in using a more "objective" approach but, perhaps more significantly, they have frequently employed a multidimensional approach to evaluating programs. Especially at the undergraduate level, a range of criteria--and multiple indicators of those criteria--have been used to assess program quality. Substantively, studies at the graduate level have yielded defensible rankings of "leading" departments in terms of faculty scholarly achievement--rankings which generally have a close correspondence to rankings in the most recent ACE-sponsored reputational studies (Cartter, 1966; Roose and Anderson, 1970). At the undergraduate level, "top" programs have been identified--though the wide difference in rankings across studies indicates considerable disagreement over the top-ranked programs.

In summary, this overall approach to program evaluation has considerable promise. If researchers can more effectively respond to the methodological and substantive criticisms raised here, and above all seek to anchor their criteria in theory and relate them to program goals, then research in this tradition can truly become more "objective." Program evaluation research based on objective indicators is needed at the graduate, professional, and undergraduate levels.

Research on Quantitative Correlates of Quality

In the last few years, a growing number of researchers have searched for quantitative correlates of quality. Put simply, researchers have attempted to identify objective program characteristics, labeled "correlates of quality," that are associated with programs that are taken to be of high quality.

Research on the quantitative correlates of quality has followed a similar general format. The researcher first selects one or more fields that have been evaluated and ranked in a reputational study (in some studies, more than one field is examined so that comparisons can be made across fields). Then, by either comparing simple (bivariate) correlations of one or more variables, or by using multiple regression or multiple discriminate function analysis to estimate the relationship between groups of variables and those programs having high quality ratings, the researcher distills a cluster of objective traits associated with program quality.

Most studies in this tradition have examined the correlates of program quality as measured in the reputational studies by Cartter (1966) and Roose-Anderson (1970). Since the two ACE studies focused exclusively on evaluating graduate program quality, most researchers have examined the correlates of program quality solely at the graduate level. However, one major study has been completed at the undergraduate level (Astin and Solomon, 1981; Solomon and Astin, 1981) and it is reviewed briefly after a more extended treatment of the

literature on the correlates of graduate program quality.

Cartter (1966:112-114) was the first to find strong relationships between several objective correlates--including faculty compensation and library resources--and institutional rankings. Shortly thereafter, a number of other researchers followed his lead, examining the relationship between a range of potential correlates of quality and the Cartter rankings in specific disciplines.

Many early studies focused on graduate programs in sociology. For example, Lewis (1968) found some correspondence between measures of faculty and student publication and the Cartter rankings of sociology departments. Other investigators found relationships between the Cartter report rankings of sociology departments and various indicators of quality, including individual prestige (Lightfield, 1971), research productivity (Knudsen and Vaughan, 1969; Glenn and Villemez, 1970), student-faculty ratios (Janes, 1969; Lavendar, Mathers, and Pease, 1971), and hiring patterns (Gross, 1970; Schichor, 1970). Studies conducted later, both in sociology and other single disciplines, identified such correlates as size (Elton and Rogers, 1971; Elton and Rose, 1972) and research productivity (Drew, 1975; Guba and Clark, 1978) as well as a number of other correlates (Abbott, 1972; Solomon, 1972; Solomon and Walters, 1976).

In the last decade, several researchers have worked with subsets of the ACE-rated departments. Hagstrom (1971), for example, sampled 125 departments in four scientific fields: physics, chemistry, mathematics, and biology. He found large, significant correlations between departmental prestige (Cartter ratings of the quality of graduate faculty) and department size, research production, research opportunities, faculty background, student characteristics (including undergraduate selectivity), and faculty awards and offices. Subsequent studies which sampled at least four major fields also found relation-

ships between the ACE ratings and other quality indicators: a National Science Board (1969) study found fourteen correlates of quality; Beyer and Snipper (1974) identified fourteen independent variables; Morgan, Kearney, and Regens (1976) found four predictor variables; and Drew and Karpf (1981) identified a powerful single predictor variable--departmental rate of publication in highly cited journals.

While the search for empirical correlates of quality has yielded some impressive findings, several caveats are in order. First, many of the correlates that researchers have identified are highly interrelated with one another, and the independent effects of specific correlates have not always been isolated. Second, the magnitude of correlations varies somewhat across comparable studies, due to the particular correlates and measures of those correlates that researchers have used. Third, and relatedly, some researchers have found some noteworthy differences between the relative importance of some key correlates across different fields (Hagstrom, 1971; Beyer and Snipper, 1974).

Although the magnitude of correlations varies somewhat across studies and disciplines, they have been generally high enough to conclude that researchers have been quite successful in identifying quantitative correlates of departmental quality at the graduate level (as indicated in the ACE reputational studies) (Lawrence and Green, 1980:24). As Hagstrom stated a decade ago: "Unless it is possible to devise some systematic causal theory, it will make little sense to seek additional predictors of departmental prestige or productivity. More refined measurements will permit only slight improvements in the proportion of variance explained" (1971: 389). Summarized in Appendix B are the major correlates of departmental quality at the graduate level that have been identified by researchers in the studies cited above.

While research on the quantitative correlates of quality has focused almost

exclusively on the graduate level, a recent study by Astin and Solomon (1981) examined objective correlates of "reputational" quality ratings in seven undergraduate fields. In brief, Astin and Solomon wanted to know if it was possible to estimate undergraduate quality by examining objective, quantifiable information about an institution.

While the researchers found some differences across fields, they identified a number of factors that correlated quite highly with their reputational ratings of "overall quality of undergraduate education" (Solomon and Astin, 1981). The major correlates they identified included: size, prestige, selectivity, per student financial expenditures (such as expenditures for educational and general purposes), and measures of curricular concentration (for example, higher quality was associated with relatively large concentration of bachelor's degrees in the natural sciences and engineering). Significantly, they found that a combination of prestige (weighted positively) and size (weighted negatively) provided a highly accurate estimate of overall quality ratings in each of the seven fields.

The study by Astin and Solomon is an important piece of research for a number of substantive and methodological reasons, not least of which is that it focuses attention on quality at the undergraduate level. Yet as Astin and Solomon themselves suggest, more research is needed on the correlates of quality at the undergraduate level. In particular, research is needed which is based on assessments of quality other than the Solomon-Astin (1981) ratings and which uses quality ratings from a larger sample of institutions and undergraduate fields.

Critique

Based on this discussion of the research on quantitative correlates, it is tempting to conclude that this approach has been highly successful--at

least in the graduate domain. After all, researchers seem to have reached considerable agreement on the major correlates of quality at the graduate level, and some scholars have concluded that further research at the graduate level is unlikely to make much of a contribution (cf. Hagstrom, 1971:389). In our view, however, this conclusion is premature, and we begin this critique with a discussion of the major limitation of research in this tradition.

First of all, almost all of the research has examined correlates of quality as measured by the ACE-sponsored reputational studies of graduate programs at leading institutions (Cartter, 1966; Roose and Anderson, 1970), now more than a decade old. Significantly, researchers have not searched for correlates of program quality as measured by objective indicators or through such studies as accreditation reports and statewide reviews of program quality. As a consequence of being completely dependent on the ACE ratings of program quality, research on the quantitative correlates of quality is, in turn, subject to all of the limitations associated with these reputational studies of program quality.

Because of this dependency of correlational studies on the ACE-sponsored reputational studies, the main limitations of those studies bear repeating: they have methodological limitations (such as rater bias); they focus exclusively on the graduate (doctoral) level; they measure program quality based largely on faculty reputation; and they focus only on the historically recognized institutions. Even if the methodological limitations are minimized, the singular fact remains that quality has been identified only at the "top" institutions, at one level (doctoral), using one major criterion (faculty quality). Thus, because of their exclusive reliance on reputational studies for the identification of "quality" programs, correlational studies are limited in terms of their generalizability.

Further research is needed which overcomes the limitations associated with

exclusive reliance on reputational ratings to identify program quality. More specifically, researchers should identify high quality programs at institutions which have been rated on criteria in addition to (and other than) faculty quality, and which are not limited to the "top-ranked" ACE institutions. Either by conducting their own studies or by relying on various independent assessments of program quality, researchers should use ratings of program quality that are not as limited in scope and generalizability as the ACE studies.

Second, most studies have used an "atheoretical" approach in identifying quantitative correlates of quality. That is, instead of identifying potential correlates on the basis of a theory of quality (which, in turn, would suggest specific variables that may be associated with quality), many investigators seem to have "rummaged" through their data in search of any and all factors that might conceivably be empirically linked to program quality--even if they have little or no defensible relationship to quality. (Exceptions include Drew and Karpf, 1981 and Solomon and Astin, 1981). Thus, even though extant research has yielded considerable agreement on the major correlates of program quality at the graduate level, there are altogether too few theoretical explanations as to why particular correlates make conceptual sense. If we are to move toward the development of some theory about quality, the selection of correlates should be linked to some emerging theory of quality. Otherwise, what we are left with is a series of empirical correlates of quality that lack any theoretical justification.

Third, many of the correlates of quality that have been identified are highly interrelated with one another (for example, size and prestige), and many researchers have failed to isolate the independent effects of individual correlates. In some cases, this omission is due to limitations associated with the research design. For example, studies that rely on bivariate cor-

relations cannot, by definition, isolate the independent effects of individual correlates. In other cases, researchers using multivariate approaches (such as regression analysis) have not examined the interrelatedness of variables.

Last, studies of the objective correlates of quality have focused almost exclusively on the graduate domain. Aside from the recent study discussed above (Solomon and Astin, 1981), little attention has been given to the undergraduate level and no studies have been conducted using professional schools.

As noted earlier, there have been some impressive findings regarding the quantitative correlates of graduate program quality. Although researchers may be justly criticized for often failing to link their findings to any theoretical scaffolding, the fact remains that a number of correlates have been identified that seem to have a strong association with program quality. Accordingly, we are persuaded that this approach to studying program quality has rich possibilities. At the undergraduate and professional levels, as well as the graduate level, further research is needed which addresses the methodological criticisms raised above and, equally important, is based on evaluations of program quality that are not encumbered by the limitations of the ACE reputational studies. The ACE studies have been a convenient measure of program quality, but other assessments of program quality are needed if the major correlates identified are to be linked with a more multifaceted conception of quality. If future research is based on these suggestions, this approach may prove the most fruitful of all in the ongoing search for a better understanding of "quality" in American higher education.

REFERENCES

Abbott, W.F. "University and Departmental Determinants of the Prestige of Sociology Departments." American Sociologist 7 (November, 1972): 14-15.

Adams, Arvil V. and Krislov, Joseph. "Evaluating the Quality of American Universities: A New Approach." Research in Higher Education 8 (1978): 97-109.

American Education Council, in "Survey of 37 Academic Fields and 130 Schools in U.S., Finds Graduate Facilities Have Improved Substantially Between '64 and '69 at Most Schools." New York Times, January 3, 1971, p. 54.

Astin, Alexander W. and Henson, James W. "New Measures of College Selectivity." Research in Higher Education 6 (1977): 1-8.

_____ and Solomon, Lewis C. "Measuring Academic Quality: An Interim Report." Change 11 (1979): 48-51.

_____. "Are Reputational Ratings Needed to Measure Quality?" Change 13 (October, 1981): 14-19.

Barak, Robert J. "A Survey and Analysis of State-Level Academic Program Review in Higher Education." Unpublished Doctoral Dissertation, State University of New York at Buffalo, 1976.

Beyer, Janice M. and Snipper, Reuben. "Objective Versus Subjective Indicators of Quality." Sociology of Education 47 (1974): 541-57.

Blackburn, Robert T. and Lingenfelter, Paul E. Assessing Quality in Doctoral Programs: Criteria and Correlates of Excellence. Ann Arbor: Center for the Study of Higher Education, University of Michigan, 1973.

Blau, Peter M. and Margulies, Rebecca Z. "The Reputations of American Professional Schools." Change 6 (Winter, 1974-75): 42-47.

Bowker, Albert H. "Quality and Quantity in Higher Education." Journal of the American Statistical Association (March, 1964): 1-15.

Brown, David G. The Mobile Professors. Washington, D.C.: American Council on Education, 1967.

Calvert, Jack G.; Pitts, James N., Jr.; and Dorion, George H. Graduate School in the Sciences: Entrance, Survival, and Careers. New York: Wiley-Interscience, 1971.

Carpenter, Ray L. and Carpenter, Patricia A. "The Doctorate in Librarianship and an Assessment of Graduate Library Education." Journal of Education for Librarianship 11 (1970): 3-45.

Cartter, Allan A. An Assessment of Quality in Graduate Education. Washington, D.C.: American Council on Education, 1966.

Cartter, Allan A. and Solomon, Lewis C. "The Cartter Report on the Leading Schools of Education, Law and Business." Change 9 (1977): 44-48.

Clark, Mary Jo. "The Meaning of Quality in Graduate and Professional Education." In Scholars in the Making: The Development of Professional Students, edited by Joseph Katz and Rodney T. Hartnett. Cambridge, Mass.: Ballinger, 1976.

_____; Hartnett, Rodney T.; and Baird, Leonard L. Assessing Dimensions of Quality in Doctoral Education: A Technical Report of a National Study in Three Fields. Princeton, N.J.: Educational Testing Service, 1976.

Clemente, Frank and Sturgis, Richard B. "Quality of Department of Doctoral Training and Research Productivity." Sociology of Education 47 (1974): 287-99.

Clewell, Beatriz C. "Assessing Educational Quality in Higher Education." Unpublished Doctoral Dissertation, The Florida State University, 1980.

Cole, Jonathan R. and Lipton, James A. "The Reputation of American Medical Schools." Social Forces 55 (1977): 662-84.

Cox, W. Miles and Catt, Viola. "Productivity Ratings of Graduate Programs in Psychology Based on Publication in the Journals of the American Psychological Association." American Psychologist 32 (1977): 793-813.

Drew, David E. Science Development: An Evaluation Study. Washington, D.C.: National Academy of Sciences, 1975.

_____ and Karpf, Ronald. "Ranking Academic Departments: Empirical Findings and a Theoretical Perspective." Research in Higher Education 14 (1981): 305-320.

Dubé, W.F. "Undergraduate Origins of U.S. Medical Students." Journal of Medical Education 49 (1974): 1005-10.

Eells, Walter C. "Leading American Graduate Schools, 1948-1958." Liberal Education 46 (March, 1960): 16-20.

Elton, Charles F. and Rogers, Samuel A. "Physics Department Ratings: Another Evaluation." Science 174 (1971): 565-68.

_____ and Rose, H.A. "What are Ratings Rating?" American Psychologist 27 (1972): 197-201.

"The 15 Top-Ranked Business Schools in the United States." MBA (December, 1974): 21-25.

Fotheringham, David J. "Quality Indicators in Undergraduate Education as Viewed by Selected Respondents in Selected Colleges and Universities." Unpublished Doctoral Dissertation, State University of New York at Albany, 1978.

Glenn, Norval D. and Villemez, Wayne. "The Productivity of Sociologists at 45 American Universities." American Sociologist 5 (1970): 224-52.

Gourman, Jack. The Gourman Report: Ratings of American Colleges. Phoenix, Arizona: Continuing Research Institute, 1967.

Guba, Egon G. and Clark, David L. "Levels of R & D Productivity in Schools of Education." Educational Researcher 7 (1978): 3-9.

Gross, George R. "The Organization Set: A Study of Sociology Departments." American Sociologist 5 (February, 1970): 25-29.

Hagstrom, Warren O. "Inputs, Outputs and the Prestige of University Science Departments." Sociology of Education 44 (1971): 375-97.

Hartnett, Rodney T.; Clark, Mary Jo; and Baird, Leonard L. "Reputational Ratings of Doctoral Programs." Science 199 (1978): 1310-14.

Hughes, Raymond M. "Report of the Committee on Graduate Instruction." Educational Record 15 (1934): 192-234.

_____. A Study of the Graduate Schools in America. Oxford, Ohio: Miami University Press, 1925.

Hurlbert, Beverly M. "Status and Exchange in the Profession of Anthropology." American Anthropologist 78 (1976): 272-284.

Janes, Robert W. "The Student Faculty Ratio in Graduate Programs of Selected Departments of Sociology." American Sociologist 4 (May, 1969): 123-27.

Johnson, Richard R. "Leadership Among American Colleges." Change 10 (November, 1978): 50-51.

Jordan, Robert T. "Library Characteristics of Colleges Ranking High in Academic Excellence." College and Research Libraries 24 (1963): 369-76.

Kelso, Charles D. "How Does Your Law School Measure Up?" Student Lawyer 4 (1975): 20-24.

Keniston, Hayward. Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania. Philadelphia: University of Pennsylvania Press, 1959.

Knudsen, Dean D. and Vaughan, Ted R. "Quality in Graduate Education: A Re-evaluation of the Rankings of Sociology Departments in the Cartter Report." American Sociologist 4 (1969): 12-19.

Krause, Ervin D. and Krause, Loretta. "The Colleges that Produce our Best Scientists: A Study of the Academic Training Grounds of a Large Group of Distinguished American Scientists." Science Education 54 (1970): 133-40.

Lavendar, Abraham D.; Mathers, Richard A.; and Pease, John. "The Student-Faculty Ratio in Graduate Programs of Selected Departments of Sociology: A Supplement to the Janes Report." American Sociologist 6 (February, 1971): 29-30.

Lawrence, Judith K. and Green, Kenneth C. A Question of Quality: The Higher Education Ratings Game. AAHE-ERIC/Higher Education Research Report No. 5, Washington, D.C.: American Association for Higher Education, 1980.

Lewis, Lionel S. "On Subjective and Objective Rankings of Sociology Departments." American Sociologist 3 (1968): 129-31.

Lightfield, Timothy E. "Output and Recognition of Sociologists." American Sociologist 6 (May, 1971): 128-33.

Longanecker, David A. "Assessing Quality in Graduate Education: A Multivariate Approach." Unpublished Doctoral Dissertation, Stanford University, 1978.

Lynd, Albert B. "The Identification of Criteria for Evaluating Graduate Programs in Higher Education at Big Ten Institutions." Unpublished Doctoral Dissertation, Michigan State University, 1976.

Magoun, H.W. "The Cartter Report on Quality in Graduate Education." Journal of Higher Education 37 (1966): 481-92.

Margulies, Rebecca Z. and Blau, Peter. "America's Leading Professional Schools." Change 5 (1973): 21-27.

Morgan, David R.; Kearney, Richard C.; and Regens, James L. "Assessing Quality among Graduate Institutions of Higher Education in the United States." Social Science Quarterly 57 (1976): 671-79.

Munson, C. Eric and Nelson, Pat. "Measuring the Quality of Professional Schools." UCLA Educator 19 (1977): 41-52.

National Science Board. Graduate Education: Parameters for Public Policy. Washington, D.C.: National Science Foundation, 1969.

Petrowski, William R.; Brown, Evan L.; and Duffy, John A. "National Universities and the ACE Ratings." Journal of Higher Education 44 (1973): 495-513.

"The Popular Vote: Rankings of Top Schools." Juris Doctor 6 (1976): 17-18, 21.

Roose, Kenneth D. and Anderson, Charles J. A Rating of Graduate Programs. Washington, D.C.: American Council on Education, 1970.

Scully, Malcolm G. "The Well-Known Universities Lead in Rating of Faculties' Reputations." Chronicle of Higher Education, January 15, 1979, pp. 6-7.

Shichor, David. "Prestige of Sociology Departments and the Placing of New Ph.D.'s." American Sociologist 5 (1970): 157-60.

Siebring, B. Richard. "The American Council on Education Rankings of Quality in Graduate Education and Membership in the National Academy of Science." Science Education 53 (1969): 73-77.

Smith, Richard and Fiedler, Fred E. "The Measurement of Scholarly Work: A Critical Review of the Literature." Educational Record 52 (1971): 225-32.

Solomon, Lewis C. and Astin, Alexander W. "Departments Without Distinguished Graduate Programs." Change 13 (September, 1981): 23-28.

Solomon, W.E. "Correlates of Prestige Rankings of Graduate Programs in Sociology Departments." American Sociologist 7 (November, 1972): 14-15.

_____ and Walters, A.T. "The Relationship Between Productivity and Prestige of Graduate Sociology Departments: Fact or Artifact?" American Sociologist 10 (1976): 229-36.

Somit, Alfred and Tanenhaus, Joseph. American Political Science: A Profile of a Discipline. New York: Atherton Press, 1964.

Tidball, M. Elizabeth and Kistiakowski, V. "Baccalaureate Origins of American Scientists and Scholars." Science 193 (1976): 646-52.

"The Top 15: The MBA Survey of Graduate Business Schools." MBA (December, 1975): 33-35.

Webster, David S. "Advantages and Disadvantages of Methods of Assessing Quality." Change 13 (October, 1981): 20-24.

ADDITIONAL REFERENCES

Astin, Alexander W. Four Critical Years. San Francisco: Jossey-Bass, 1977.

Bowen, Howard R. Investment in Learning. San Francisco: Jossey-Bass, 1977.

Juris Doctor. "The Popular Vote: Rankings of Top Schools." 6 (1967): 17-18,21.

MFA. "The 15 Top-Ranked Business Schools in the United States." (December, 1974): 21-25.

MBA. "The Top 15: The MBA Survey of Graduate Business Schools." (December, 1975): 33-35.